2

AD-A221 747

Approximate Inference and Scientific Method

Mark Fulk and Sanjay Jain

Technical Report 313
October 1989

UNIVERSITY OF
ROCHESTER
COMPUTER SCIENCE

Approximate Inference and Scientific Method

Mark Fulk [1] and Sanjay Jain [2]
Department of Computer Science
University of Rochester
Rochester, New York 14627

October 1989

[1]Supported by ONR/DARPA research contract number N00014-82-K-0193 and NSF grant CCR 8320136 to the University of Rochester. email address fulk@cs.rochester.edu

[2]Supported by NSF grant CCR 8320136 to the University of Rochester. email address jain@cs.rochester.edu

Abstract

A new identification criterion, motivated by notions of successively improving approximations in the philosophy of science, is defined. It is shown that the class of recursive functions is identifiable under this criterion. This result is extended to permit somewhat more realistic types of data than usual. This criterion is then modified to consider restrictions on the quality of approximations.

1 Introduction

Research in inductive inference has historically been motivated by considerations of the philosophy of science, e.g. [CS83]. However, the criteria of success so far proposed seem unrealistic for science.

In the sequel we assume that scientific experiments and observations are encoded as natural numbers and that the process of scientific theory formation can be modeled by an algorithmic device which operates on the encoded experiments and observations.

Gold's [Gol67] criterion demands that an inductive inference machine produce a final correct program (in the sense that it correctly computes the input function or set); others [CS83] have liberalized that criterion to allow final programs that are correct except on finitely many inputs. [Bar74] and [CS83] also give criteria that permit infinite sequences of programs, nearly all of which are (perhaps only nearly) correct.

We hold with Peirce [Pei58] that science cannot be expected to produce a final theory of anything, nor even a cofinal sequence of nearly correct theories. Instead, the best we can hope for is that science produces an infinite sequence of improving approximations to reality. It is the point of this paper to give precision to this notion.

2 Notations

$\mathcal{N}$ is the set of natural numbers. i, j, l, m, n, x, y, z and variously decorated versions thereof range over natural numbers unless otherwise specified. d ranges over the real interval $[0..1]$. f, g range over functions from $\mathcal{N}$ to $\mathcal{N}$. $\emptyset$ denotes the null set. $Card(S)$ denotes the cardinality of the set S. max, min denote the maximum and minimum of a set respectively. $\mu x[Q(x)]$ is the least natural number x such that $Q(x)$ is true (if such exists).

$\mathcal{R}$ denotes the class of recursive functions. φ denotes an acceptable numbering [Rog58], [Rog67]. Φ denotes an arbitrary Blum complexity measure [Blu67] for φ. $\langle \cdot, \cdot \rangle$ stands for an arbitrary computable one to one encoding of all pairs of natural numbers onto $\mathcal{N}$ [Rog67]. π_1 and π_2 are the corresponding projection functions. $\langle \cdot, \cdot \rangle$ is extended to n-tuples in the usual way. For any two partial functions η_1 and η_2, $\eta_1 =^n \eta_2$ means that $\text{card}(\{x | \eta_1(x) \neq \eta_2(x)\}) \leq n$. $\eta_1 =^* \eta_2$ means that $\text{card}(\{x | \eta_1(x) \neq \eta_2(x)\})$ is finite. For any two sets S_1 and S_2, $S_1 =^n S_2$ denotes $\text{card}((S_1 - S_2) \cup (S_2 - S_1)) \leq n$. $S_1 =^* S_2$ denotes

card$((S_1 - S_2) \cup (S_2 - S_1))$ is finite. $S_1 \oplus S_2$ denotes the symmetric difference of the sets S_1 and S_2.

Sequences are functions with domain an initial segment of $\mathcal{N}$. An information sequence is an infinite sequence; a segment is a finite sequence. t ranges over information sequences. $\sigma, \sigma_0, ..$ range over segments. t_n denotes the initial segment of t with length n. content$(t) = range(t) - \{*\}$; intuitively it is a set of meaningful things presented in t. Similarly, $content(\sigma) = range(\sigma) - \{*\}$. $length(\sigma)$ denotes the length of σ. $\sigma_1 \diamond \sigma_2$ denotes the concatenation of σ_1 and σ_2. For all recursive functions f, $f|^n$ denotes the finite segment $((\langle 0, f(0)\rangle), (\langle 1, f(1)\rangle), (\langle 2, f(2)\rangle), ...(\langle n, f(n)\rangle))$.

$\mathbf{M}, \mathbf{M}_0, \ldots$ denote IIMs. $\mathbf{M}(\sigma)$ is the last output of $\mathbf{M}$ after receiving input σ (note that σ can be encoded as a natural number). $\mathbf{M}(t) \downarrow = i$ iff $(\overset{\infty}{\forall} n)[\mathbf{M}(\overline{t_n}) = i]$. We write $\mathbf{M}(t)\downarrow$ iff $(\exists i)[\mathbf{M}(t)\downarrow = i]$. Any unexplained notation is from [Rog67].

3 Preliminaries

In this section we briefly discuss notions from recursion theoretic machine learning literature. For detailed discussion see [OSW86; CS83; Gol67; AS83; KW80; BB75].

Inductive Inference machines (IIMs) have been used in the study of identification of recursive functions as well as recursively enumerable languages [Gol67; CS83; CL82; AS83; KW80; Ful85; BB75]. For function learning the input sequence given to the **IIM** is $\langle 0, f(0)\rangle, \langle 1, f(1)\rangle, \ldots$, where f is the function being learned. A criterion of success (called **Ex**-identification) is for the machine to eventually output a last program, which computes (nearly computes) f. Formally,

Definition 1 [Gol67; BB75; CS83] $\mathbf{M}$ $\mathbf{Ex}^a$*-identifies* f (written $f \in \mathbf{Ex}^a(\mathbf{M})$) iff both $\mathbf{M}(f)\downarrow$ and $\varphi_{\mathbf{M}(f)} =^a f$.

Definition 2 [Gol67; BB75; CS83] $\mathbf{Ex}^a = \{\mathcal{S} \subseteq \mathcal{R} | (\exists \mathbf{M})[\mathcal{S} \subseteq \mathbf{Ex}^a(\mathbf{M})]\}$.

In the above definitions a stands for the number of anomalies allowed in the final program. $a = *$ means that unbounded but finite number of anomalies is allowed in the final program. Case and Smith [CS83] introduced another infinite hierarchy of identification

criterion which we describe below. "**BC**" stands for *behaviorally correct.* Barzdin [Bar74] independently introduced a similar notion.

Definition 3 [CS83] M $\mathbf{BC}^a$-identifies f (written: $f \in \mathbf{BC}^a(\mathrm{M})$) iff, M fed f outputs over time an infinite sequence of programs $p_0, p_1, p_2, \ldots$ such that $(\overset{\infty}{\forall} n)[\varphi_{p_n} =^a f]$.

Definition 4 [CS83] $\mathbf{BC}^a = \{\mathcal{S} \subseteq \mathcal{R} | (\exists \mathrm{M})[\mathcal{S} \subseteq \mathbf{BC}^a(\mathrm{M})]\}$.

We usually write **Ex** for $\mathbf{Ex}^0$, **TxtEx** for $\mathbf{TxtEx}^0$, **BC** for $\mathbf{BC}^0$, and **TxtBC** for $\mathbf{TxtBC}^0$.

4 Approximation of recursive functions

Definition 5 M **Ap**-*identifies* f (written: $f \in \mathbf{Ap}(\mathrm{M})$) iff there is a sequence of sets $S_n \subset \mathcal{N}$ such that:

i) for all n, $\mathrm{M}(f_n)$ correctly computes f on all $x \in S_n$;

ii) for all n, $S_n \subseteq S_{n+1}$;

iii) for all x there is n with $x \in S_n$;

iv) there are infinitely many n with $S_{n+1} - S_n$ infinite.

If M **Ap**-identifies f, then we also say that M *approximates* f.

Theorem 1 *There is an inductive inference machine* M *that approximates every recursive function.*

Proof: The idea is that M partitions $\mathcal{N}$ into infinitely many infinite subsets; M then carries out a separate induction process for each subset. In the inference process for one subset, M uses the number of the subset as a bound on the Gödel number of programs to be considered.

When $\mathrm{M}(f|^n)$ is run on an input $x \geq n$ from the i-th element of the partition, it uses the program less than i that best fits $f|^n$, where n is used to bound the computation of the degree of fit. For inputs $x < n$, $\mathrm{M}(f|^n)$ outputs $f(x)$, which is in the input.

Let patch, select, err, and best be recursive functions such that:

$$\varphi_{\text{patch}(i,\sigma)}(x) = \begin{cases} y, & \text{if } (\exists y)\langle x, y\rangle \in \text{content}(\sigma), \\ \varphi_i(x) & \text{otherwise;} \end{cases}$$

$$\varphi_{\text{select}(n,\langle a_0,\ldots,a_n\rangle)}(\langle i, x\rangle) = \begin{cases} \varphi_{a_i}(\langle i, x\rangle), & \text{if } i < n; \\ \varphi_{a_n}(\langle i, x\rangle), & \text{otherwise;} \end{cases}$$

$$\text{err}(j,\sigma) = \mu x[x = |\sigma| \text{ or } \Phi_j(x) > |\sigma| \text{ or } \varphi_j(x) \neq \sigma(x)];$$

$$\text{best}(b,\sigma) = \mu i.[i \leq b \wedge (\forall j \leq b)[err(i,\sigma) \geq err(j,\sigma)]].$$

patch patches σ into program i. select chooses one of $a_0, \ldots, a_n$ and runs it according to the left projection of the input. err finds the first apparent error committed by j relative to σ. best finds the best program (as measured by *err*) for σ and less than b.

Let $\mathbf{M}(\sigma)$ be $patch(\sigma, select(|\sigma|, \langle best(0,\sigma), \ldots, best(|\sigma|,\sigma)\rangle))$.

Note that for any sufficiently large i and n, $best(i, f \mid^n)$ will be a program for f. We may take $S_n = \{0, \ldots, n-1\} \bigcup_{(k \mid k<n \wedge (\forall n' \geq n)\varphi_{\text{best}(i,n')}=f)}(\{\langle k, x\rangle \mid x \in \mathcal{N}\})$. Verification of the properties (i-iv) in Definition (5) is immediate. □

5 Density Restrictions on Ap-criterion

The algorithm given in the previous section **Ap**-identifies all the recursive functions. It is easy to choose pairing function such that for all recursive function f the limiting density of the sets S_n is positive. In this section we study the effects of requiring the limiting density of the sets $S_0, S_1, \ldots$ so formed to be above a certain prespecified value. First we formally define what we mean by density of a set.

Definition 6 [Roy86] The *density* of a set $A \subseteq \mathcal{N}$ in a finite and nonempty set B (denoted: $\mathbf{d}(A;B)$) is $\text{card}(A \bigcap B)/\text{card}(B)$.

Intuitively, $\mathbf{d}(A;B)$ can be thought of as the probability of selecting an element of A when choosing an arbitrary element from B.

Definition 7 [Roy86] The *density* of a set $A \subseteq \mathcal{N}$ (denoted: $\mathbf{d}(A)$) is $\lim_{n\to\infty} \inf\{\text{d}(A; \{z \mid z \leq x\}) \mid x \geq n\}$.

Definition 8 Let $d \in [0,1]$. An **IIM M** $\mathbf{DAp}^d$*-identifies* a function f (written: $f \in \mathbf{DAp}^d(\mathbf{M})$) iff there exists a sequence of sets $S_n \subset \mathcal{N}$ such that:

(i) for all n, $\mathbf{M}(f_n)$ correctly computes f on all $x \in S_n$;

(ii) for all n, $S_n \subseteq S_{n+1}$;

(iii) for all x there is n with $x \in S_n$;

(iv) there are infinitely many n with $S_{n+1} - S_n$ infinite;

(v) $lim_{n \to \infty} \mathbf{d}(S_n) \geq d$.

Definition 9 Let $d \in [0,1]$. $\mathbf{DAp}^d = \{\mathcal{C} \subseteq \mathcal{R} \mid (\exists \mathbf{M})[\mathcal{C} \subseteq \mathbf{DAp}^d(\mathbf{M})]\}$

Even though the limiting density of a set may be 1, there may be arbitrarily large gaps. We thus introduce another form of identification which prohibits such large gaps.

Definition 10 [Roy86] The *uniform density* of a set A in intervals of length $\geq n$ (denoted: $\mathbf{ud}_n(A)$) is $\inf(\{d(A;\{z \mid x \leq z \leq y\}) \mid x, y \in \mathcal{N} \textit{ and } y - x \geq n\})$. *Uniform density* of A (denoted: $\mathbf{ud}(A)$) is $\lim_{n \to \infty} \mathbf{ud}_n(A)$.

Definition 11 Let $d \in [0,1]$. An **IIM M** $\mathbf{UDAp}^d$*-identifies* a function f (written: $f \in \mathbf{UDAp}^d(\mathbf{M})$) iff there exists a sequence of sets $S_n \subset \mathcal{N}$ such that:

(i) for all n, $\mathbf{M}(f_n)$ correctly computes f on all $x \in S_n$;

(ii) for all n, $S_n \subseteq S_{n+1}$;

(iii) for all x there is n with $x \in S_n$;

(iv) there are infinitely many n with $S_{n+1} - S_n$ infinite;

(v) $lim_{n \to \infty} \mathbf{ud}(S_n) \geq d$.

Definition 12 Let $d \in [0,1]$. $\mathbf{UDAp}^d = \{\mathcal{C} \subseteq \mathcal{R} \mid (\exists \mathbf{M})[\mathcal{C} \subseteq \mathbf{UDAp}^d(\mathbf{M})]\}$

The following theorem shows that no machine can identify all of the recursive functions if the sequence of sets S_n is required to have limiting density $\geq \epsilon > 0$.

Theorem 2 $(\forall d, 0 < d \leq 1)[\mathcal{R} \notin \mathbf{DAp}^d]$

Proof: We prove the theorem for $d > 1/3$. Proof can be generalized to the case when $d > 1/n, n \in \mathcal{N}$. Suppose by way of contradiction that IIM M $\mathbf{DAp}^{1/3+\epsilon}$-identifies $\mathcal{R}$, $\epsilon > 0$. Then by ORT[Cas74] there exists a recursive 1-1 p, such that the following holds. Let φ_e^s denote φ_e defined before stage s. Let $\varphi_{p(0)}(0) = 0$. Let x_s denote the maximum x such that $\varphi_{p(0)}(x)$ is defined before stage s. Go to stage 1.

<u>Stage s</u>

Dovetail steps 1 and 2.

If 1 succeeds go to step 3.

1. Search for

1a. $l, m, l > x_s, m > l/(\epsilon/100)$;

1b. $y_{x_s+1}, y_{x_s+2}, \ldots, y_l$, and

1c. $w_1, w_2, \ldots w_k, l < w_i < m, k > m \cdot (2/3 + \epsilon/2)$

such that

$(\forall i, 1 \leq i \leq k)[\mathrm{M}(\varphi_{p(0)}|^{x_s} \diamond \langle x_s + 1, y_{x_s+1} \rangle \diamond \cdots \diamond \langle l, y_l \rangle)(w_i) \downarrow]$

2. Let $\varphi_{p(s)}(x) = \varphi_{p(0)}(x)$ for $x \leq x_s$.

Let $x_{s,s'}$ denote the maximum x such that $\varphi_{p(s)}(x)$ is defined before substage s'.

Go to substage 0.

substage s'

2.1 Search for

2.1a. $l, m, l > x_{s,s'}, m > l/(\epsilon/100)$;

2.1b. $y_{x_{s,s'}+1}, y_{x_{s,s'}+2}, \ldots, y_l$, and

2.1c. $w_1, w_2, \ldots w_k, l < w_i < m, k > m \cdot (1/3 + \epsilon/2)$

such that

$(\forall i, 1 \leq i \leq k)[\mathrm{M}(\varphi_{p(s)}|^{x_{s,s'}} \diamond \langle x_{s,s'} + 1, y_{x_{s,s'}+1} \rangle \diamond \cdots \diamond \langle l, y_l \rangle)(w_i) \downarrow]$

2.2. If and when such $l, m \cdots$ are found let,

2.2.a $\varphi_{p(s)}(x) = y_x$ for $x_{s,s'} < x \leq l$,

2.2.b Let $e = \mathrm{M}(\varphi_{p(s)}|^{x_{s,s'}} \diamond \langle x_{s,s'} + 1, y_{x_{s,s'}+1} \rangle \diamond \cdots \diamond \langle l, y_l \rangle)$;

Let $\varphi_{p(s)}(x) = \varphi_e(x) + 1$ for $x \in \{w_i | 0 < i \leq k\}$.

2.2.c Let $\varphi_{p(s)}(x) = 0$ if $x < m$ and $\varphi_{p(s)}(x)$ has not been defined till now.

Go to substage $s' + 1$

End substage s'.

3. If and when 1 succeds let $l, m \cdots$ be as found in 1. Let,

3.1.a $\varphi_{p(0)}(x) = y_x$ for $x_s < x \leq l$,

3.1.b Let $e = \mathbf{M}(\varphi_{p(0)}|^{x_s} \diamond \langle x_s + 1, y_{x_s+1}\rangle \diamond \cdots \diamond \langle l, y_l\rangle)$;

Let $\varphi_{p(0)}(x) = \varphi_e(x) + 1$ for $x \in \{w_i | 0 < i \leq k\}$.

3.1.c Let $\varphi_{p(0)}(x) = 0$ if $x < m$ and $\varphi_{p(0)}(x)$ has not been defined till now.

Go to stage $s + 1$

End stage s.

Now consider the following cases:

case 1: There are infinitely many stages.

In this case let $f = \varphi_{p(0)}$. Clearly f is recursive. We claim that no $S_0, S_1, \ldots$ can exist satisfying (i-v) in definition (8) for f and $d = 1/3 + \epsilon$. Suppose otherwise. Then there exists n_1, n_2 such that $\mathbf{d}(S_{n_1}) > 1/3 + 60/100 \cdot \epsilon$. Also for all $x > n_2$, $\mathbf{d}(S_{n_1} : [0..x]) > 1/3 + 55/100\epsilon$. But then in all the stages greater than $max(n_1, n_2)$ due to step 3 and the way $l, m, w_1, \cdots$ were chosen we have that there exists an error point for $\mathbf{M}(f|^l)$ in S_l (since $l > max(n_1, n_2)$ and the fraction of points upto m on which $\mathbf{M}(f|^l)$ commits error is at least $2/3 + \epsilon/2$). This contradicts (i) in definition (8). Thus $\mathbf{M}$ does not $\mathbf{DAp}^{1/3+\epsilon}$-identify f.

case 2: Stage s never halts. In this case, $\mathbf{M}$ does not output, a program which has a domain of limiting density more than $2/3 + 51/100 \cdot \epsilon$, on any extension of $\varphi_{p(0)}$.

case 2.1: In stage s there are infinitely many substages.

In this case let $f = \varphi_{p(s)}$. Since on f $\mathbf{M}$ never outputs a program which has a domain of limiting density more than $2/3 + 51/100 \cdot \epsilon$, arguing as in case 1 we have that $\mathbf{M}$ does not $\mathbf{DAp}^{1/3+\epsilon}$-identify f.

case 2.2: In stage s stage s' never halts. In this case on no extension of $\varphi_{p(s)}|^{x_{s,s'}}$ does $\mathbf{M}$ output a program which has domain of limiting density $1/3 + 51/100\epsilon$. Let f be any recursive function which is an extension of $\varphi_{p(s)}|^{x_{s,s'}}$. Then $\mathbf{M}$ does not $\mathbf{DAp}^{1/3+\epsilon}$-identify f.

From the above cases we have that $\mathbf{M}$ does not $\mathbf{DAp}^{1/3+\epsilon}$-identify $\mathcal{R}$.

When $d > 1/n$ the above proof can be generalized by taking $n - 1$ levels of iteration instead of 2 as done in the above procedure. We leave the details to the reader. $\Box$.

Theorem 3 $(\forall d_1, d_2 \mid 0 \le d_1 < d_2 \le 1)[\mathbf{UDAp}^{d_1} - \mathbf{DAp}^{d_2} \ne \emptyset]$

Proof: Without loss of generality let $d_1 = m/n$ and $d_2 = (m+1)/n$. Let

$\mathcal{C} = \{f \in \mathcal{R} | x < m \ \mathbf{mod}\ n \Rightarrow f(x) = 0\}$. An easy modification of the procedure to **Ap** identify all the recursive functions given in the previous section, gives us a procedure to **Ap**-identify $\mathcal{C}$ with $S_0 = \{x \mid x < m \ \mathbf{mod}\ n\}$. Thus $\mathcal{C} \in \mathbf{UDAp}^{d_1}$. Suppose by way of contradiction that **IIM M** $\mathbf{DAp}^{d_2}$-identifies $\mathcal{C}$. Then **M** can be easily modified to $\mathbf{DAp}^{1/(n-m)}$ identify all the recursive functions, contradicting Theorem (2). □

Theorem 4 $(\forall d, 0 < d \le 1)[\mathbf{DAp}^1 - \mathbf{UDAp}^d \ne \emptyset]$

Proof: Let $\mathcal{C} = \{f \in \mathcal{R} | (\forall x)(\exists n \mid 2^n < x < 2^{n+1} - n)[f(x) = 0]\}$. Again an easy modification of the procedure to **Ap** identify all the recursive functions given in the previous section, gives us a procedure to **Ap**-identify $\mathcal{C}$ with $S_0 = \{x \mid (\exists n \mid 2^n < x < 2^{n+1} - n)\}$. Thus $\mathcal{C} \in \mathbf{DAp}^1$. Suppose by way of contradiction that **IIM M** $\mathbf{UDAp}^d$-identifies $\mathcal{C}$. Then **M** can be easily modified to $\mathbf{UDAp}^d$ identify all the recursive functions, contradicting Theorem (2). □

Now we show that even though all the recursive function cannot be $\mathbf{DAp}^d$-identified, there are large classes of functions which can be $\mathbf{DAp}^1$-identified.

Theorem 5 $(\forall i \in \mathcal{N})[\mathbf{BC}^i \subseteq \mathbf{DAp}^1]$

Proof: Given **IIM M** we construct another machine $\mathbf{M}'$, which $\mathbf{DAp}^1$-identifies all functions $\mathbf{BC}^i$ identified by **M**. Let $fl(n) = \max(\{2^k : 2^k \le n\})$. Let $\mathbf{M}'(f|^n) = \mathrm{patch}(\mathbf{M}(f|^{fl(n)}), f|^n)$, where

$$\varphi_{\mathrm{patch}(i,\sigma)}(x) = \begin{cases} y, & \text{if } (\exists y)\langle x, y\rangle \in \mathrm{content}(\sigma), \\ \varphi_i(x) & \text{otherwise:} \end{cases}$$

Let n_0 be so large that for all $n \ge 2^{n_0}$, $\mathbf{M}(f|^n) =^i f$. Let $errset = \{x | (\exists n \mid n_0 \le n \le \lfloor \log x \rfloor)[\varphi_{\mathbf{M}(f|^{2^n})}(x) \ne f(x)]\}$. Clearly, if $\mathbf{M}(f|^n), n > 2^{n_0}$ commits an error on x then $x \in errset$. Also the limiting density of $errset$ is 0. Theorem follows. □

6 Further considerations and open problems

Note that **M** in the proof of Theorem (1) is clearly immune to corrected noise; as long as **M** is eventually apprised of the correct value for any given data point, it will still succeed in approximating the input function.

Scientific experiments are not always deterministic; even if quantum mechanical indeterminacy is not important, one can never be certain that all of the significant variables have been controlled. On the other hand it is arguable that the set of possible outcomes of an experiment is always finite.

These considerations lead to a formulation of inductive inference in which the function to be learned carries experimental descriptions to finite sets of outcomes, and the data to the inductive inference machine consists of experiments paired with one outcome at a time. It should be clear that, even under these circumstances, the result of Theorem (1) holds.

In some respects, these results are not very satisfactory. One would like to be able to give some account of the variations in the degree of confidence we have in the outcomes of various experiments. Also, the lower numbered partitions of the data are never completely predicted. Science does partition experiments into classes, and treat each class to some extent separately; but the classes are not simply the arbitrary choices made by a pairing function but reflect, to some extent, the results of the experiments.

It remains open whether or not one can approximate a set from positive data only; We conjecture not.

7 Acknowledgements

We would like to thank Prof. John Case, Prof. David Haussler, Arun Sharma, Rajeev Raman and Lata Narayanan for helpful discussions. Part of this work was in the dissertation of the first author at SUNY/Buffalo.

References

[AS83] D. Angluin and C. Smith. A survey of inductive inference: Theory and methods. *Computing Surveys*, 15:237–289, 1983.

[Bar74] J. A. Barzdin. Two theorems on the limiting synthesis of functions. *Latv. Gos. Univ. Uce. Zap.*, 210:82–88, 1974.

[BB75] L. Blum and M. Blum. Toward a mathematical theory of inductive inference. *Information and Control*, 28:125–155, 1975.

[Blu67] M. Blum. A machine independent theory of the complexity of recursive functions. *Journal of the ACM*, 14:322–336, 1967.

[Cas74] J. Case. Periodicity in generations of automata. *Mathematical Systems Theory*. 8:15–32, 1974.

[CL82] J. Case and C. Lynes. Machine inductive inference and language identification. *Lecture Notes in Computer Science, Springer-Verlag. Berlin*, 140. 1982.

[CS83] J. Case and C. Smith. Comparision of identification criteria for machine inductive inference. *Theoretical Computer Science*, 25:193–220, 1983.

[Ful85] M. Fulk. *A Study of Inductive Inference machines*. PhD thesis, SUNY/ Buffalo, 1985.

[Gol67] E.M. Gold. Language identification in the limit. *Information and Control*. 10:447–474, 1967.

[KW80] R. Klette and R. Wiehagen. Research in the theory of inductive inference by gdr mathematicians–a survey. *Information Sciences*, 22:149–169, 1980.

[OSW86] D. Osherson, M. Stob, and S. Weinstein. *Systems that Learn, An Introduction to Learning Theory for Cognitive and Computer Scientists*. MIT Press, Cambridge Mass., 1986.

[Pei58] C.S. Peirce. In A.W.Burks, editor, *Collected Papers*. Harvard University Press,Cambridge Mass., 1958.

[Rog58] H. Rogers. Godel numberings of partial recursive functions. *Journal of Symbolic Logic*, 23:331–341, 1958.

[Rog67] H. Rogers. *Theory of Recursive Functions and Effective Computability*. McGraw Hill, New York, 1967.

[Roy86] J. Royer. Inductive inference of approximations. *Information and Control*, 70:156–178, 1986.